Using Sketching in Machine Learning Pipeline

Dimensionality Reduction

CS-430 Project

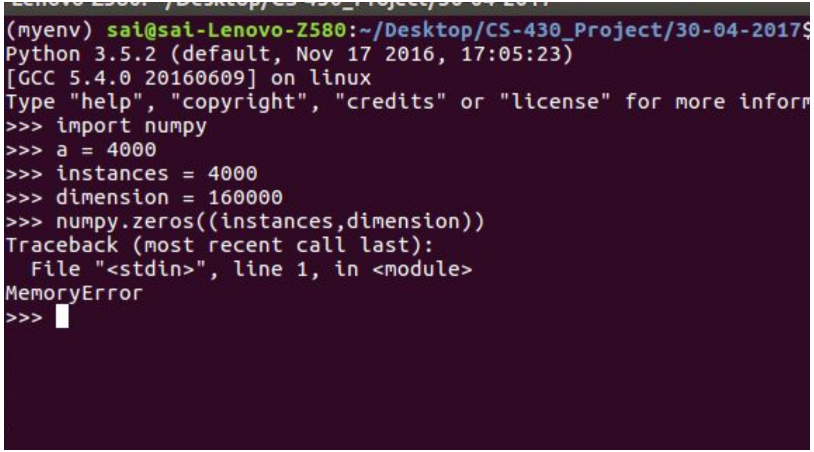
Students: Akhil Pabbathi , Kapil Pathak & Sai Chenchala

Theme:

Effect of different sketching techniques on the classifier accuracy and performance.

Motivation

- Working with very large and high dimensional data is challenging
- Resources and Time



An Elegant Solution:

Many dimensionality reduction techniques such as Feature hashing, Random Projections are available

PIPELINE

- Load Data
- Apply appropriate dimensionality reduction techniques
- Use it for machine learning

Outline

- Techniques : b bit minwise hashing , Feature hashing, JL Random Projection
- Classifiers: Linear SVM, Nonlinear SVM, Logistic Regression & Decision Tree
- Primary Data Set: Reuters Corpus Volume -1
- Secondary Data Set: Farm - Ads

A Brief Introduction - b bit minwise hashing

Computing similarity between two sets S1 and S2

$$R = |S1 \cap S2| / |S1 \cup S2|$$

Probabilistic Approach - Apply random permutations to elements of S1 and S2

$$R = Pr(\min(\Pi(S1)) = \min(\Pi(S2)))$$

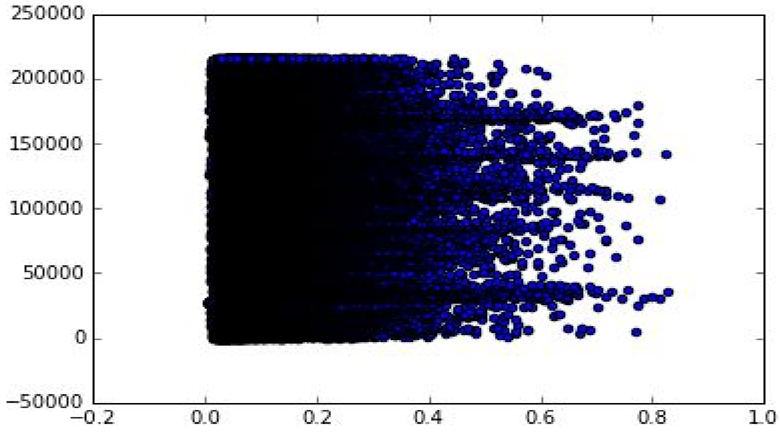
Store only the lowest b bits

Binary Quantization - Challenge

- Naive Approach: Using a threshold
- Representing N (<= 2^b) using

2^b length vector Ex: {3} = {0,0,0,1}

- Our Approach: 891 = {(0,0,0,0,0,0,0,0,1,0),(0,0,0,0,0,0,0,0,0,1),(0,1,0,0,0,0,0,0,0,0)}



Feature Hashing (Hashing trick)

- Feature hashing is mainly used for reducing the dimensionality of feature vectors while feeding them to classifiers.
- Most of the text categorization datasets are sparse in nature. Feature hashing is useful to eliminate this sparsity and save the memory space.
- It allows easy handling of missing data
- It’s not possible achieve inverse mapping

Feature Hashing (Hashing trick)

- The effect of hash collisions can be alleviated by using signed hash function

- Mathematically it can be written as,

$$\phi_i^{(h,\xi)}(x) = \sum_{j:h(j)=i} \xi(j) x_j$$

- The results are derived using MurmurHash3 hash function inbuilt in scikit learn library
- Feature hashing is used to reduce the dimensionality of each dataset to 40000, 30000 and 10000 respectively

A brief introduction on JL Random Projections

$X \in R^{nxd}$ is the data matrix with n samples in R^d

$P \in R^{dxr}$ is a random projection matrix where r<<d

The new projected data matrix would be $XP \in R^{nxr}$

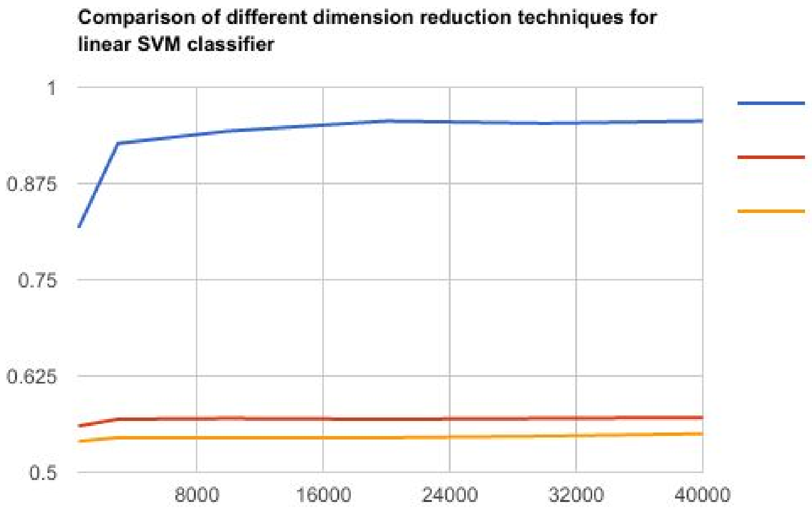
If P is carefully chosen, then all pairwise Euclidean distances are preserved with high probability.

There many possible constructions for P, one of them is a matrix whose entries are i.i.d standard Gaussian random variables.

An approach for matrix multiplication

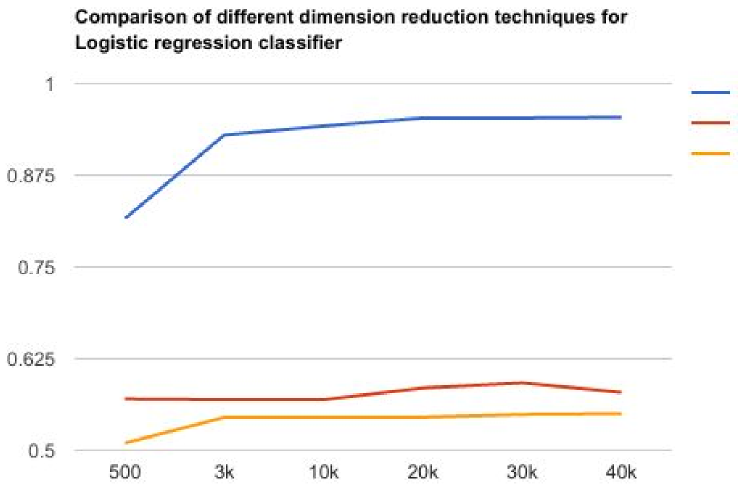
As the random projection matrix P is quite large we have divided the columns of P into a set of blocks say (P_1,P_2,....,P_{10}), then multiplied X with each of the P_i's and concatenated the results of the multiplications and created the final projected data matrix which has been used for training and testing the classifiers.

Accuracy Plot Linear SVM



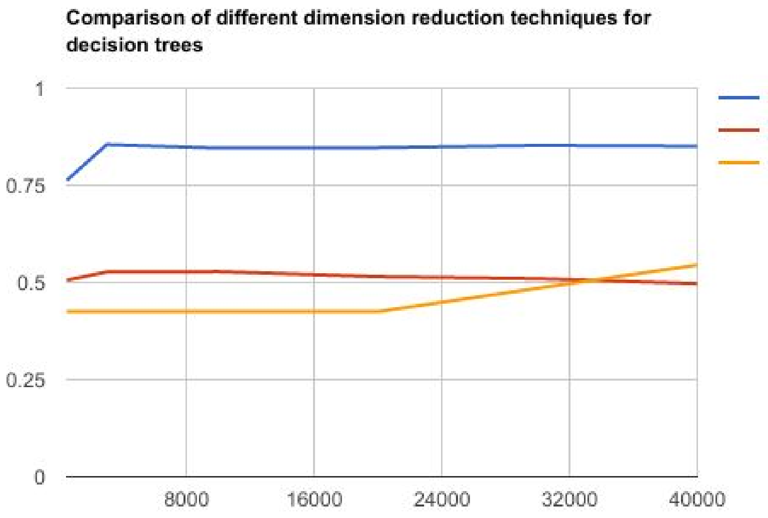
-Feature Hashing
-JL Random Projection
- b bit min hash

Accuracy Plot Linear Regression



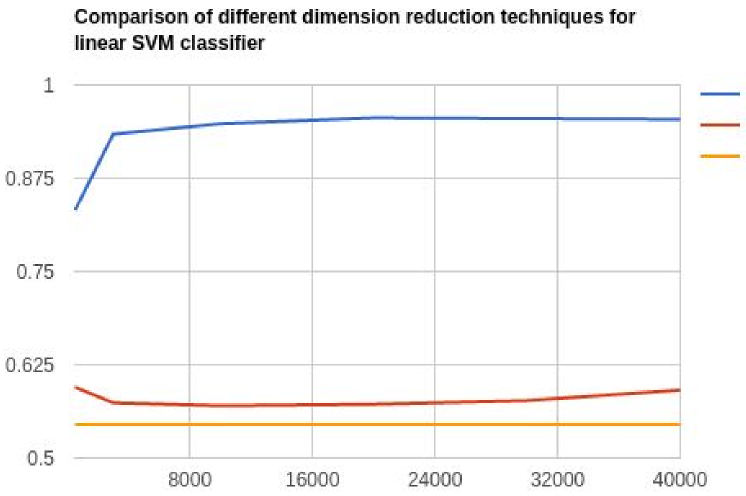
-Feature Hashing
-JL Random Projection
- b bit min hash

Accuracy Plot Decision Tree



-Feature Hashing
-JL Random Projection
- b bit min hash

Accuracy Plot Nonlinear SVM



-Feature Hashing
-JL Random Projection
- b bit min hash

References

- K. Weinberger, A. Dasgupta, J. Langford, A. Smola, and J. Attenberg. Feature hashing for large scale multitask learning. In ICML'09
- RCV-1
- Farm-Ads
- Feature Hashing Video Tutorial
- S. Paul, C. Boutsidis, M. Magdon-Ismail, and P. Drineas. Random projections for support vector machines. In Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS) (2013), pp. 498–506.
- https://arxiv.org/abs/1105.4385


THANK
YOU!
ctrl
alt